**Batch: A4 Roll No.: 16010122147**

**Experiment No 2**

**Group No: 03**

|  |
| --- |
| **Title: Literature Survey** |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Objective:** The objective of a literature survey is to review, analyze, and synthesize existing research to identify gaps, trends, and insights that inform and support a study's context and direction.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Expected Outcome of Experiment:**

|  |  |
| --- | --- |
|  | **At the end of successful completion of the course the student will be able to** |
| CO1 | Define the problem statement and scope of problem |
| CO5 | Prepare a technical report based on the Mini project. |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Books/ Journals/ Websites referred:**

**1.**

**2.**

**3.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**The students are expected to prepare chapter no 2 in the format given below**

**Chapter 2**

**Literature Survey**

*The Objective of a literature survey is to review existing research, identify gaps, and establish a strong foundation for the study. It helps in understanding key concepts, comparing different approaches, and justifying the need for the current research by analyzing past studies.*

#### ****1****. Introduction

#### AI-based image generation has become a significant breakthrough in the field of artificial intelligence, combining natural language processing (NLP) and computer vision. With advancements in generative models, particularly diffusion models like ****Stable Diffusion 2.1**** and ****3.5 Large****, it has become possible to produce high-resolution, detailed images from text prompts. This literature survey aims to explore existing methodologies in text-to-image synthesis, analyze prior works on diffusion models, and identify the need for improved control, realism, and performance in AI image generation. The goal is to establish a foundation that supports the objectives of the current mini project, “NeuraPix – AI Image Generator. 2. Review of Existing Literature

Several generative AI models have evolved in recent years, with notable contributions in the areas of latent diffusion, hierarchical image generation, and prompt-to-image accuracy. Stability AI’s diffusion models and OpenAI’s CLIP-based techniques have greatly enhanced the quality and control of generated images.

* **Rombach et al. (2022)** introduced Latent Diffusion Models (LDM) that reduced computational cost while maintaining high image fidelity.
* **Ramesh et al. (2022)** proposed a hierarchical text-conditional image generation framework using CLIP latents, which demonstrated effective prompt understanding.
* **Dhariwal and Nichol (2021)** demonstrated how diffusion models outperform GANs on image synthesis tasks, introducing structured denoising processes.
* **Saharia et al. (2022)** focused on photorealistic image generation using text prompts with deep semantic understanding, contributing to diffusion model optimization.
* **Ho et al. (2020)** laid the foundation of diffusion probabilistic models, setting the stage for subsequent advances in generative AI.

**3. Related Work**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Paper Title (Including Author Details, Year of publication, Conference/Journal | Methodology | Dataset Used | Observation of proposed methodology | Pros | Cons | Findings |
| **"High-resolution Image Synthesis with Latent Diffusion Models"** – Rombach et al., CVPR 2022 | Used latent diffusion in compressed space to generate images efficiently | COCO, OpenImages | Reduced compute while maintaining high quality | Low memory, high resolution | May blur fine details | Enabled scalable image generation with fewer resources |
| **Hierarchical Text-Conditional Image Generation with CLIP Latents"** – Ramesh et al., arXiv 2022 | Used CLIP and transformers to improve prompt-image alignment | Internal OpenAI datasets | Highly accurate text-to-image results | Excellent semantic alignment | Requires heavy compute | Foundation of DALL·E 2 |
| **Diffusion Models Beat GANs on Image Synthesis"** – Dhariwal & Nichol, NeurIPS 2021 | Improved denoising diffusion model with classifier guidance | CIFAR-10, ImageNet | Outperformed GANs in image quality | Stable, interpretable | Slower than GANs | Proved diffusion as a superior generative method |
| **Photorealistic Text-to-Image Diffusion with Deep Language Understanding"** – Saharia et al., arXiv 2022 | Combined image diffusion with deep NLP models | LAION-400M | High realism and better prompt adherence | Rich language understanding | Potential bias in training | Strong baseline for photorealistic generation |
| **Denoising Diffusion Probabilistic Models"** – Ho et al., NeurIPS 2020 | Introduced basic DDPM framework for generative tasks | CIFAR-10 | Foundation for all diffusion-based models | Simple, effective | High training cost | Core model behind newer diffusion models |
| **"ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models"** – Lvmin Zhang et al., 2023 | Enabled edge, pose, depth-based control in generation | MS-COCO + custom controls | Greatly improved control over output images | Fine control, multimodal | Complexity in control input | Advanced interactive image creation |
| **"GLIDE: Towards Photorealistic Image Generation and Editing with Text-guided Diffusion Models"** – Nichol et al., 2022 | Introduced editing + generation using diffusion + guidance | Public image-text pairs | Enabled image manipulation via prompts | Text-based editing | Limited resolution | Allowed creative flexibility |
| **"Imagen: Photorealistic Text-to-Image Diffusion Models with Large Language Models"** – Saharia et al., 2022 | Combined LLMs with diffusion for detailed generation | Internal datasets | Higher photorealism than DALL·E | |  | | --- | | Deep NLP + image quality |  |  | | --- | |  | | Not open-source | Major milestone in realistic image generation |
| **"Versatile Diffusion: Text, Images and Beyond"** – Kim et al., 2023 | Unified multi-modal inputs (text, sketches) for generation | MS-COCO, LAION | Handles multiple input types | Flexible interface | Still under research | Useful for creative tools |
| **“Stable Diffusion: High-resolution Image Synthesis using Latent Text-to-Image Diffusion"** – Stability AI, 2022 | Combines U-Net + CLIP + latent space diffusion | LAION-5B | Open-source, highly customizable | Fast inference, public availability | May lack high semantic depth | Enabled community-driven generative tools |

#### 4. Research Gaps and Challenges

* **Gaps in Fine Control**: Although diffusion models generate high-quality images, controlling specific aspects of image generation (pose, depth, edges) remains a challenge.
* **Computational Cost**: Advanced models such as Stable Diffusion 3.5 require substantial computational resources for real-time or large-scale deployment.
* **Prompt Adherence Issues**: Despite advances, exact interpretation of complex or abstract prompts still requires fine-tuning.
* **Ethical Concerns**: Potential misuse of generated content for deepfakes or biased outputs raises critical ethical issues.
* **Future Direction**: More precise control mechanisms like ControlNets (Blur, Canny, Depth) and improved multimodal integration (text, sketch, image input) could address current limitations.